Available online at www.sciencedirect.com
SCIENCE DIRECT
Remote Sensing of Environment 95 (2005) 290–302
Remote Sensing of Environment
www.elsevier.com/locate/rse

Signature extension through space for northern landcover classification: A comparison of radiometric correction methods

Ian Olthof[a,*], Chris Butson[b], Robert Fraser[c]

[a]Noetix Research Inc., Canada
[b]Prologic Resraech Inc., Canada
[c]Natural Resources Canada, Environmental Monitoring Section, Applications Division, Canada Centre for Remote Sensing, 588 Booth St., Ottawa, Canada K1A 0Y7

Received 7 October 2004; received in revised form 12 December 2004; accepted 12 December 2004

Abstract

Northern landcover mapping for climate change and carbon modeling requires greater detail than what is available from coarse resolution data. Mapping landcover with medium resolution data from Landsat presents challenges due to differences in time and space between scene acquisitions required for full coverage. These differences cause landcover signatures to vary due to haze, solar geometry and phenology, among other factors. One way to circumvent this problem is to have an image interpreter classify each scene independently, however, this is not an optimal solution in the north due to a lack of spatially extensive reference data and resources required to label scenes individually. Another possible approach is to stabilize signatures in space and time so that they may be extracted from one scene and extended to others, thereby reducing the amount of reference data and user input required for mapping large areas. A radiometric normalization approach was developed that exploits the high temporal frequency with which coarse resolution data are acquired and the high spatial frequency of medium resolution data. The current paper compares this radiometric correction methodology with an established absolute calibration methodology for signature extension for landcover classification and explores factors that affect extension performance to recommend how and when signature extension can be applied. Overall, the new normalization method produced better extension and classification results than absolute calibration. Results also showed that extension performance was affected more by geographical distance than by differences in anniversary dates between acquisitions for the range of data examined. Geographical distance in the north–south direction leads to poorer extension performance than distance in the east–west direction due in part to differences in vegetation composition assigned the same class label in the latitudinal direction. While extension performance was somewhat variable and in some cases did not produce a best classification result by itself, it provided an initial best guess of landcover that can subsequently be refined by an expert image interpreter.

Keywords: Signature extension; Radiometric correction; northern landcover

1. Introduction

The need for a detailed landcover map of northern Canada motivated the development of an automated Landsat normalization and compositing procedure in Olthof et al. (2005). This method uses surface reflectance from coarse resolution SPOT VEGETATION (VGT) imagery as a reference for inter-scene Landsat normalization, and is expected to produce more consistent radiometry than either absolute calibration or other relative normalization methods. The technique was subsequently employed to create a 250-m resampled Landsat coverage of northern Canada (Olthof et al., submitted for publication). The full-resolution scenes that were used in this mosaic and their normalization coefficients can be applied to create seamless 30-m regional composites. Mosaicing the entire database to achieve a complete northern coverage with the full resolution data is currently not feasible due to the prohibitively large file size that would result.

* Corresponding author. Tel.: +1 613 947 1233; fax: +1 613 947 1406.
E-mail address: iolthof@ccrs.nrcan.gc.ca (I. Olthof).

0034-4257/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2004.12.015

Landcover information is nonetheless required at resolutions finer than 250 m. For example, the United Nations Framework Convention on Climate Change (UNFCCC) requires each participating country to produce a national inventory of carbon stocks in 1990 and anthropogenic greenhouse gas emissions in subsequent reporting periods from all sources, including land-use change. Most land use such as roads and smaller settlements cannot be detected in 250-m to 1-km resolution imagery. Furthermore, land uses that can be detected in such coarse resolution imagery cannot be delineated precisely to allow an accurate assessment of their area. Exclusion of fine-scale land-use change from national carbon stock inventories would result in a gross underestimation of carbon loss due to such land-use changes as new road construction. The need to include fine-scale information for carbon accounting requires data from a medium resolution sensor such as Landsat. Because all Landsat scenes in Canada's national database cannot be classified simultaneously, methods are needed to classify each scene (or tiles of scenes) separately and in a consistent manner to provide the information required for UNFCCC reporting. The extension of spectral signatures in space and time to classify each scene independently is one approach to accomplish this.

Signature extension is widely applied in mapping with remote sensing data, as it is the basis of supervised classification where signatures are extracted from a known feature and used to classify the same feature type located at some distance in time and space from the signature source. However, in the context of mapping with Landsat data, supervised classification is usually performed within a single scene or a mosaic consisting of a number of radiometrically normalized scenes. In the current paper, we investigate the potential to extend signatures between Landsat scenes after radiometric correction for large area mapping with a large number (>100) of individual Landsat scenes.

Signature extension was first investigated for the Large Area Crop Inventory Experiment (LACIE) in the mid-1970s to predict crop type and yield from Landsat MSS imagery (Minter, 1978). Results obtained in that study were poor, though extension in time produced encouraging results if haze was the only source of signature variation, while extension in space degraded classifier performance and was not correctable using sun angle and haze corrections. Although signature extension continued to be investigated for geological (Clark, 1990) and, to a lesser extent, vegetation mapping (Bojinski et al., 2003) using spectral libraries and hyperspectral sensors, it quietly disappeared from the landcover mapping literature with Landsat data as most analyses were confined to local regions. Currently, larger regional to national level databases are being assembled for large landcover mapping initiatives (Franklin & Wulder, 2002) and have generated renewed interest in signature extension, though most efforts have been modest thus far. Woodcock et al. (2001) investigated signature extension using neural networks for discriminating forest change from no change, and Pax-Lenney et al. (2001) applied similar methods to differentiate conifer from non-conifer forest.

Signature extension performance is dependent on a feature's radiometric consistency through time and space. The poor performance of most previous attempts to extend signatures is primarily a function of inadequate radiometric calibration or normalization between scenes. Pax-Lenney et al. (2001) examined the effects of different calibration methods on signature extension and found that simple dark object subtraction (DOS) techniques performed as well as more complex radiative transfer approaches. The authors also highlighted the need for better operational corrections for signature extension at regional scales.

Radiometric consistency can be improved through normalization; however, certain factors that affect signature extension will persist after normalization since they cannot easily be corrected. A scene's time of acquisition affects solar geometry, which can produce a different spectral response depending on surface illumination and roughness. Acquisition day of year can lead to similar variation in illumination and also produces phenological differences of vegetated surfaces. The spectral properties of a thematic class can vary with geographical distance when assigning the same label to a slightly different vegetation association. While radiometric normalization can help to minimize some of these effects, it cannot remove them altogether.

The normalization technique developed in Olthof et al. (2005) should provide more consistent normalization of distant scenes than DOS since all scenes are normalized to a common reference consisting of coarse resolution composite data. The temporal frequency with which coarse resolution data are acquired and advanced compositing techniques allow selection of the clearest pixel within a growing season that most closely represents surface reflectance. This normalization approach exploits the best features from both coarse and medium resolution data by using the high temporal frequency and wide area coverage of coarse resolution and high spatial frequency of medium resolution satellite data.

Other large area landcover mapping initiatives using Landsat data have relied on a scene-by-scene classification approach. For example, Vogelmann et al. (2001) performed individual scene classification to produce landcover information for all of the US at 30-m spatial resolution. Wulder et al. (2003) are using a similar approach to map Canada's forested areas. While highly labour intensive, this approach is feasible in both cases due to spatially extensive and accurate reference data available for each scene in the respective areas of interest. In remote areas such as Canada's North, adequate reference data and other resources are currently not available for a scene-by-scene classification approach. Signature extension provides an opportunity to extrapolate expert knowledge embedded in a classification to areas where reference data do not exist.

This paper will investigate factors that affect signature extension for independently mapping landcover of a large number of individual Landsat scenes after normalization to VGT to help demonstrate its potential and limitations. As a benchmark test, an accepted atmospheric correction methodology using Dark Dense Vegetation (DDV) (Kaufman et al., 1997) is also applied prior to signature extension. Within each of the two Landsat correction methods, signature extension in space and time will be examined and compared between methods.

2. Methods

Ten images from the SILC (Satellite Information for Landcover of Canada) database (Table 1) that formed part of the normalized northern mosaic were used as reference landcover in this study (Fig. 1). The SILC database currently consists of 33 individual Landsat-scene classifications providing information on the spatial and structural characteristics of the landcover of Canada. SILC has been used previously as reference data for accuracy assessment and scaling of coarse resolution landcover products (Cihlar et al., 2003; Latifovic & Olthof, 2004).

The 10 classified Landsat ETM+images selected for this study represented the same acquisition (same scene, date and time) as images contained in the normalized northern mosaic, thereby avoiding errors due to phenology and atmosphere. However, images input to SILC and those in the northern database were georeferenced independently, and it was noticed that small georeferencing errors existed in the SILC classifications. SILC was produced before images became available through the Centre for Topographic Information (CTI) and were georeferenced using National Topographic Data Base (NTDB) vector coverages. The images were subsequently orthorectified by CTI when they were made available to the public through Geogratis (http://geogratis.cgdi.gc.ca/). Each SILC classification was registered to the images obtained from CTI and therefore some georeferencing error may still have existed in the SILC classification, adding a small amount of error in the assessment of all classifications using signature extension.

Table 1
SILC scenes

Scene (path, row)	Date	LCC scene centre	Landcover association
18022	8/20/2000	1218722, 7381651	South
20018	8/2/2000	1043969, 7962973	Intermediate
21014	7/11/2001	981486, 8568174	North
22023	7/15/2000	826734, 7120437	South
26011	7/27/2000	719005, 8938142	North
30021	8/8/2000	102631, 7346983	South
33021	7/10/1999	−178907, 7349452	South
36009	8/2/2000	261560, 9128863	North
44013	8/10/2000	−495703, 8585672	North
52010	8/2/2000	−667369, 9073152	North

All SILC data are georeferenced to Lambert Conformal Conic (LCC) projection with 95 d W, 0 d N as the true origin and 49 d N and 77 d N as standard parallels.

The 10 SILC classifications were grouped according to two broad landcover associations. Northern scenes consisted of few, primarily sparsely vegetated landcover classes while southern scenes consisted of a greater number of classes that included boreal and tree-line transitional classes. Signatures were extended across these landcover associations, though few classes were shared between northern and southern scenes in some cases.

2.1. SILC classifications

An unsupervised classification using Landsat bands 3, 4 and 5 was used to produce individual SILC classifications. A combined approach was employed that uses features of three procedures: K-Means clustering (Bezdek, 1973) to derived a large number of initial clusters; Classification by Progressive Generalization (CPGcs; Cihlar et al., 1998) to automatically merge these to ~70, without a significant loss of landcover information; and Enhancement-Classification Methodology (ECM; Beaubien et al., 1999) for checking the quality of these two steps, further analyst-controlled merging, and the assignment of clusters to specific landcover classes (labeling). Cihlar et al. (2002) provide a more detailed discussion of the combined procedure. The Federal Geographic Data Committee (FGDC) National Vegetation Classification System (NVCS) legend, consisting of 46 landcover classes (Table 2), was employed in the SILC classifications and this analysis.

2.2. Landsat radiometric correction methods

2.2.1. SILC absolute radiometric calibration using DDV

The dense, dark vegetation (DDV) approach to approximate a scene-dependent Aerosol Optical Depth at 550 nm (AOD550) with scene-specific geographic information was used to derive surface reflectance for each of the ten Landsat scenes. This approach extracts spectral information from selected DDV pixels across various spectral wavelengths within a 100-pixel moving window to determine the contribution of path radiance influencing the pixel radiances recorded at the satellite sensor. The 6S radiative transfer code (Vermote et al., 1997) was then used iteratively to determine an estimate of surface reflectance given the calculated top-of-atmosphere reflectance and an estimated AOD550. The method is similar to the one applied by Liang et al. (2001) with several modifications that were highlighted in Butson and Fernandes (2004). Dense, dark vegetation targets were defined from a top-of-atmosphere Landsat ETM+band 7 threshold of 0.001–0.005 and a Normalized Difference Vegetation Index (NDVI) greater than the 90th percentile of the histogram. In several cases, the algorithm could not extract any DDV targets so an AOD550 value of 0.06 was used to be representative of the

Fig. 1. Landsat ETM+scenes contained in the northern mosaic for which SILC classifications exist.

typical aerosol conditions over the Canadian landmass as suggested by Fedosejevs et al. (2000).

2.2.2. SILC radiometric normalization to SPOT VGT

SPOT data consisted of a growing season-long composite produced for the Global Landcover 2000 project (GLC 2000) (Latifovic et al., 2004). Daily SPOT data consist of 1 km at-surface apparent reflectance bands in red (0.61–0.68 um) near infrared (NIR) (0.78–0.89 um) and shortwave infrared (SWIR) (1.58–1.75 um), which have nearly identical wavelengths to Landsat bands 3, 4 and 5. These data were composited at the Canada Centre for Remote Sensing (CCRS) using VGT Manager; a software package specifically designed to process VGT to remove atmospheric effects and normalize pixels to a consistent geometry of 0 degrees viewing angle and 45° solar zenith angle. Specific details on VGT data processing are presented by Latifovic et al. (2004).

Each Landsat scene was resampled to 1-km resolution using the VGT point spread function and then randomly sampled with spatially coincident pixels in the VGT composite. These data were used to generate normalization functions using a robust regression called Thiel-Sen (Fernandes & Leblanc, in press) separately for red, NIR and SWIR bands with the Landsat data as the regressor and VGT as the response variable. Theil-Sen regression uses the median of all pairwise slopes to calculate the gain of the regression equation and applies it to each pair of (x, y) observations in the sample to generate a set of offsets. The median value of this set of offsets is taken as the maximum likelihood offset for normalization. The use of median values for the slope and offset makes this regression technique relatively insensitive to leverage from outliers in x and y that can adversely affect a non-robust regression such as least squares. The Thiel-Sen gain and offset coefficients were then applied to the full-resolution Landsat imagery on scene and band-wise bases.

2.3. Signature extraction and extension

Each of the 10 scenes was used in turn as a reference from which signatures were extracted and applied to classify all scenes in the set. In order retain maximum radiometric fidelity during normalization, Du et al. (2001) suggest applying an adjustment factor to all normalization coefficients in the set so that the smallest gain is greater or equal to one, ensuring no information loss through data range compression. This approach may not be desirable when the set of normalized scenes is large (>100), since this adjustment would invariably result in data loss by saturation of some scenes. Because the normalization procedure compressed the spectral range of certain scenes, especially in the red band, input spectral data to the SILC and signature extension classifications were not identical. Therefore, the

Table 2
FGDC legend

Tree Dominated (tree crown density>25%)	
1	Evergreen forest (>75% cover)—old
2	Evergreen forest (>75% cover)—young
3	Deciduous forest (>75% cover)
4	Mixed coniferous (50–75% coniferous)—old
5	Mixed coniferous (50–75% coniferous)—young
6	Mixed deciduous (25–50% coniferous)
7	Evergreen open canopy (40–60% cover)—moss–shrub understory
8	Evergreen open canopy (40–60% cover)—lichen–shrub understory
9	Evergreen open canopy (25–40% cover)—shrub–moss understory
10	Evergreen open canopy (25–40% cover)—lichen (rock) understory
11	Deciduous open canopy (25–60% cover)
12	Deciduous open canopy-low regenerating to young broadleaf cover
13	Mixed evergreen-deciduous open canopy (25–60% cover)
14	Mixed deciduous (25–50% coniferous trees; 25–60% cover)
15	Low regenerating to young mixed cover
Shrub dominated	
16	Deciduous shrubland (>75% cover)
Herb dominated	
17	Grassland, prairie region
18	Herb–shrub–bare cover, mostly after perturbations
19	Shrubs–herb–lichen–bare
20	Wetlands
21	Sparse coniferous (density 10–25%), shrub–herb–lichens cover
22	Sparse coniferous (density 10–25%), herb–shrub cover
23	Herb–shrub
24	Shrub–herb–lichen–bare
25	Shrub–herb–lichen–water bodies
26	Lichen–shrubs–herb, bare soil or rock outcrop
27	Lichen–shrubs–herb, bare soil/rock outcrop, water bodies
28	Low vegetation cover (bare soil, rock outcrop)
29	Low vegetation cover, with snow
30	Woodland–cropland
31	Cropland–woodland
32	Annual row-crop forbs and grasses—high biomass
33	Annual row-crop forbs and grasses—medium biomass
34	Annual row-crop forbs and grasses—low biomass
Nonvascular dominated	
35	Lichen barren
36	Lichen–shrub–herb–bare
37	Sparse coniferous (density 10–25%), lichens–shrub–herb cover
Vegetation not dominant	
38	Rock outcrop, low vegetation cover
39	Recent burns
40	Mostly bare disturbed areas (e.g. cutovers)
41	Low vegetation cover
42	Urban and built-up
43	Water bodies
44	Mixes of water and land
45	Snow/ice
46	Clouds

signature extension methodology was applied to the reference scene to serve as a benchmark test for the extension of signatures to all other scenes in the set.

The method by which signatures were extracted to classify the reference scene and extended to classify other scenes was as follows (Fig. 2):

Normalized spectral data of a reference scene from which signatures were extended were clustered to 150 clusters using a k-means classifier. A lookup table was generated to relate spectral clusters obtained from the normalized data to thematic classes in the reference SILC scene using a majority rule. Signature extension was done by sampling cluster signatures in the reference scene, transferring these signatures to all other scenes in the set and using them in a minimum distance classifier to produce similar clusters in other scenes. The lookup table from the reference image was then applied to convert clusters to thematic classes. Accuracy assessments were performed with the SILC classifications as the 'truth' for the assessment of that scene.

Not all scenes contained the same classes, and therefore null classes existed in many of the classifications. Accuracy assessments included only classes that were present in both the reference and classified scenes. Because the southern landcover association contained a significantly greater number of classes than the northern association, accuracy measures such as overall accuracy and Kappa statistic (Cohen, 1960) (Eq. (1)) could not be used by themselves to examine factors affecting signature extension between associations due to the well known dependence of accuracy on the number of classes present in the assessment. Thus, the Kappa statistics and Kappa variance (Bishop et al., 1975) (Eq. (2)) were used to calculated Kappa standard or z-scores (Z_{kappa}) (Eq. (3)) to minimize the dependence of the assessment on the degrees of freedom.

$$K = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}x_{+i}}{N^2 - \sum_{i=1}^{r} x_{i+}x_{+i}} \tag{1}$$

$$\sigma_k^2 = \frac{1}{N}\left(\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4}\right) \tag{2}$$

where

$$\theta_1 = \sum_{i=1}^{r} P_{ii}$$

$$\theta_2 = \sum_{i=1}^{r} P_{i+}P_{+i}$$

$$\theta_3 = \sum_{i=1}^{r} P_{ii}(P_{i+} + P_{+i})$$

$$\theta_4 = \sum_{i=1}^{r}\sum_{j=1}^{r} P_{ij}\left(P_{i+} + P_{+j}\right)^2$$

$$P_{ii} = \frac{x_{ii}}{N}$$

$$P_{i+} = \frac{x_{i+}}{N}$$

$$P_{+i} = \frac{x_{+i}}{N}$$

$$P_{ij} = \frac{x_{ij+}}{N}$$

$$Z_{\text{kappa}} = \frac{K}{\sigma_k} \quad (3)$$

Factors thought to affect signature extension between northern and southern landcover associations were examined using the Kappa standard scores, while factors were also examined within-association (where N was relatively constant) using the Kappa statistics.

Factors considered included geographical distance in longitudinal and latitudinal directions as well as Euclidean distance, and differences between acquisition anniversary dates. Previous investigations have reported decreasing signature extendibility with increasing Euclidean distance without consideration of extension direction. Signature extension through time has also been investigated, with increasing time between acquisitions producing poorer extension results (Pax-Lenney et al., 2001).

3. Results

3.1. Radiometric correction comparison

Separate 10-by-10 matrices were produced for each of the two radiometric correction methods, with the ten scenes from which signatures were extracted forming rows and the same 10 scenes which signatures were extended to for classification forming columns (Tables 3 and 4).

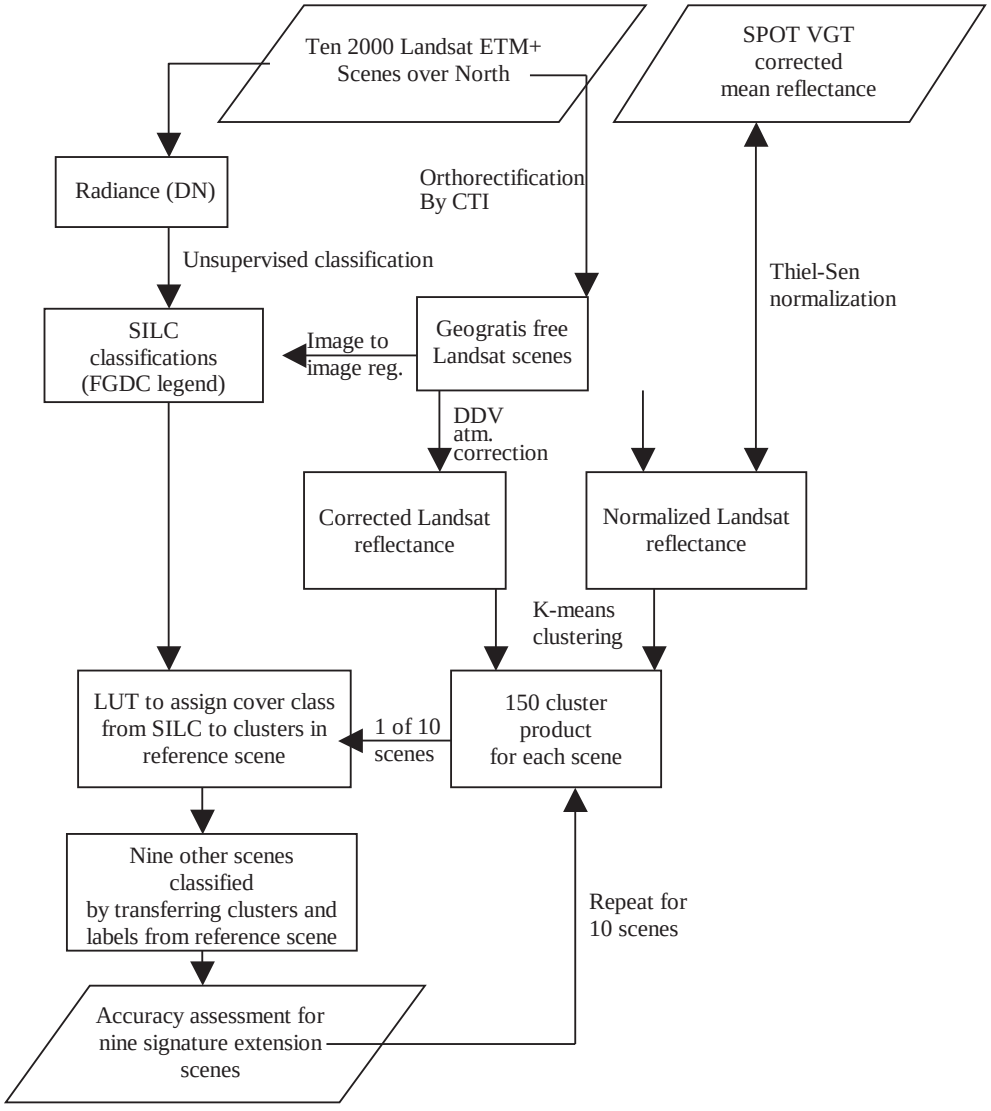

Fig. 2. Flowchart showing signature extension methodology.

Table 3
Classification accuracy (Kappa statistic and standardized Kappa (Z_{kappa}) in parentheses) of signature extension after Landsat DDV calibration and number of classes common between reference and classified scenes used in the assessments

Classification signature source (reference)	18022	20018	21014	22023	26011	30021	33021	36009	44013	52010
18022	58.6% (6434) N = 15	43.6% (4014) N = 4		30.8% (3565) N = 13	19.1% (430) N = 2	13.3% (2743) N = 10	11.2% (951) N = 12		85.7% (631) N = 2	30.1% (180) N = 3
20018	11.4% (465) N = 7	58.7% (6439) N = 11	77.1% (5099) N = 4	11.7% (251) N = 7	30.4% (1707) N = 4	0.1% (2) N = 4	2.4% (43) N = 6		67.4% (928) N = 4	32.1% (1081) N = 5
21014		82.9% (6123) N = 3	73.1% (8563) N = 6	0.0% (0) N = 2	56.1% (5429) N = 6			2.8% (76) N = 2	21.6% (1672) N = 5	42.7% (4585) N = 6
22023	33.0% (3104) N = 12	18.9% (189) N = 4		58.4% (6601) N = 19	0.0% (0) N = 2	13.7% (1852) N = 11	26.6% (2766) N = 17	89.5% (8086) N = 2	15.4% (50) N = 2	0.0% (0) N = 2
26011	98.6% (7109) N = 2	31.9% (464) N = 4	57.2% (4300) N = 6	0.0% (0) N = 2	67.9% (7610) N = 8	0.0% (0) N = 2		0.0% (0) N = 2	14.5% (878) N = 6	10.1% (917) N = 8
30021	35.9% (2100) N = 11	46.1% (3404) N = 5		20.9% (3056) N = 14	10.0% (69) N = 2	52.9% (5724) N = 17	16.3% (1926) N = 16		1.2% (10) N = 3	-0.1% (0) N = 2
33021	13.2% (1122) N = 12	47.8% (3428) N = 3		31.4% (3462) N = 18	27.4% (69) N = 2	9.2% (969) N = 15	60.9% (7178) N = 23		16.2% (63) N = 2	
36009			65.2% (5474) N = 4	1.5% (23) N = 2	49.7% (4576) N = 5			87.4% (12099) N = 5	37.6% (2296) N = 3	57.5% (4106) N = 5
44013	78.6% (1902) N = 3	71.6% (2747) N = 3	36.3% (2581) N = 4	7.8% (51) N = 2	37.6% (4097) N = 6	0.0% (0) N = 2	0.2% (3) N = 2	0.0% (0) N = 3	67.8% (6941) N = 7	44.6% (3250) N = 6
52010	22.2% (27) N = 2	49.2% (3451) N = 4	38.5% (3386) N = 6	0.0% (0) N = 2	31.4% (3966) N = 8	0.0% (0) N = 2		71.3% (7048) N = 4	34.2% (2158) N = 5	75.7% (8174) N = 8

Accuracy measures such as the Kappa statistic and Z_{kappa} as well as the number of classes common between the classified scene and the SILC classification used in the assessments are entries in the matrices. Diagonal entries represent reference classifications accuracies using signatures extracted from the same reference scene and methods described above. These accuracies represent a theoretical maximum achievable accuracy using signatures extended from other scenes. Reference scene accuracies resulting from classification using reference scene signatures are never 100% due to differences in quantization levels resulting from radiometric correction.

Tables 3 and 4 reveal similar accuracies for Landsat scenes classified using their own signatures for both Landsat normalization to VGT and absolute calibration using DDV. In Table 3, an average kappa of scenes classified using their own signatures after DDV correction (average of diagonal Kappa entries) is 66.1% and the Z_{kappa} is 7576 compared to 65.6% and 7642 for normalization to VGT in Table 4. The average overall accuracy of the 10 reference scenes was 71.8% and 71.3% for DDV correction and normalization to VGT, respectively. Differences between these are caused by different normalization coefficients from the two radiometric correction methods applied. These accuracies are also somewhat low due to a loss of radiometric fidelity in the red band caused by normalization and slightly different methods used to generate SILC classifications and classifications from signature extension.

While both radiometric correction methods produce similar accuracies for reference scenes classified using their own signatures, large differences exist when considering only classification accuracies resulting from signature extension. In Table 3, the average Kappa of classifications resulting from signature extension after absolute calibration using DDV (average of off-diagonal Kappa entries) is 25.4% and the Z_{kappa} is 1713 compared to 40.3% and 2327 for normalization to VGT in Table 4. Thus, signature extension after DDV correction achieves 38.4% of its theoretically maximum achievable kappa and 22.6% of its maximum Z_{kappa}, while normalization to VGT attains 61.4% and 30.5%, respectively.

These results demonstrate the limitations of signature extension without consideration of factors that may influence extendibility. For this comparison, signatures were extracted from each reference scene to classify itself and all

Table 4
Classification accuracy (Kappa statistic and standardized Kappa (Z_{kappa}) in parentheses) of signature extension after Landsat normalization to VGT and number of classes common between reference and classified scenes used in the assessments

Classification signature source (reference)	18022	20018	21014	22023	26011	30021	33021	36009	44013	52010
18022	57.8% (6693) $N = 15$	50.7% (2961) $N = 7$	99.9% (11159) $N = 2$	48.9% (4423) $N = 13$	89.5% (1834) $N = 2$	26.4% (2146) $N = 10$	36.4% (3367) $N = 12$		99.8% (5226) $N = 2$	30.8% (256) $N = 3$
20018	77.6% (5221) $N = 6$	56.8% (6043) $N = 12$	81.5% (4191) $N = 4$	41.4% (3591) $N = 8$	68.9% (4945) $N = 4$	94.6% (7221) $N = 4$	48.0% (2810) $N = 6$	51.1% (2810) $N = 6$	54.4% (2497) $N = 4$	78.0% (2787) $N = 5$
21014	99.6% (7301) $N = 2$	70.6% (3844) $N = 4$	76.7% (9523) $N = 7$	0.7% (11) $N = 2$	65.4% (6481) $N = 6$			78.9% (7594) $N = 4$	41.4% (3728) $N = 5$	50.1% (4950) $N = 7$
22023	30.8% (2412) $N = 12$	25.2% (855) $N = 7$		58.7% (6604) $N = 23$	0.0% (0) $N = 2$	22.6% (2509) $N = 13$	32.3% (3412) $N = 19$	37.1% (292) $N = 2$	31.5% (35) $N = 2$	11.9% (56) $N = 2$
26011	98.1% (5056) $N = 2$	48.0% (1542) $N = 4$	30.8% (2440) $N = 6$	0.0% (0) $N = 2$	72.9% (8536) $N = 8$	74.0% (316) $N = 2$		58.2% (4551) $N = 5$	20.0% (1168) $N = 6$	13.9% (1230) $N = 8$
30021	28.9% (2100) $N = 11$	71.3% (2734) $N = 5$		40.5% (4259) $N = 14$	58.7% (369) $N = 2$	53.6% (5780) $N = 17$	41.7% (4130) $N = 16$		14.8% (87) $N = 3$	2.2% (9) $N = 2$
33021	27.2% (2207) $N = 13$	35.8% (1353) $N = 6$		37.1% (4135) $N = 21$		32.5% (3365) $N = 16$	51.8% (5833) $N = 26$		82.3% (363) $N = 2$	
36009			63.5% (4575) $N = 4$	47.1% (871) $N = 2$	30.3% (2238) $N = 5$			86.6% (11623) $N = 5$	54.7% (2388) $N = 3$	59.2% (4128) $N = 5$
44013	98.6% (5752) $N = 3$	74.2% (2385) $N = 3$	33.6% (2799) $N = 5$	0.0% (0) $N = 2$	42.5% (3954) $N = 6$	2.8% (20) $N = 3$	53.3% (60) $N = 2$	54.1% (1721) $N = 3$	62.7% (5813) $N = 7$	16.5% (815) $N = 6$
52010	86.1% (2475) $N = 3$	73.1% (3417) $N = 5$	26.2% (2079) $N = 7$	2.7% (41) $N = 2$	51.8% (3906) $N = 8$	38.3% (0) $N = 2$		16.8% (756) $N = 5$	21.7% (1086) $N = 5$	78.2% (9970) $N = 9$

remaining scenes, and therefore does not represent an optimal case of signature extension. Factors affecting signature extension will now be investigated to attempt to determine conditions for a suitable application.

3.2. Factors affecting signature extension

Trends in the Z_{kappa} as a function of factors discussed in the introduction were examined excluding cases where reference scenes were classified using their own signatures. Since both northern and southern landcover associations were examined together and each association has a different number of landcover classes, standardized Kappa values (Z_{kappa}) were first used in this analysis (Fig. 3). Northern and southern landcover associations were then examined separately using only the Kappa statistic since it is a standard measure of classification accuracy and within each landcover association, all scenes contains a similar number of classes (Fig. 4).

Differences in the number of days between scene acquisitions reveal some surprising results, as both correction methods show trends of increasing accuracy with increasing number of days between acquisitions. These trends are likely accounted for by other factors, since scenes acquired more days apart have also been acquired closer together in space. Therefore, geographical distance appears to be a more important factor controlling the decrease in accuracy from signature extension than day of year for these scenes acquired near the peak of the growing season (July and August).

Examination of the effects of geographical distance on signature extension shows decreasing classification accuracy with increasing Euclidean distance between scenes. The rate of decrease is similar between radiometric correction methods, with normalization to VGT producing higher accuracies across the range of distances examined. Between 1500 and 2000 km from a signature source, classification accuracy measured by the Z_{kappa} generally drops to approximately 50% of the classification accuracy at less than 500 km from the source for both methods.

While decreasing classification accuracy with increasing distance from a signature source has been reported previously, this trend has never been examined in longitudinal and latitudinal directions to the authors' knowledge. Normalization to VGT produced only a very slight decrease in classification accuracy with increasing distance in

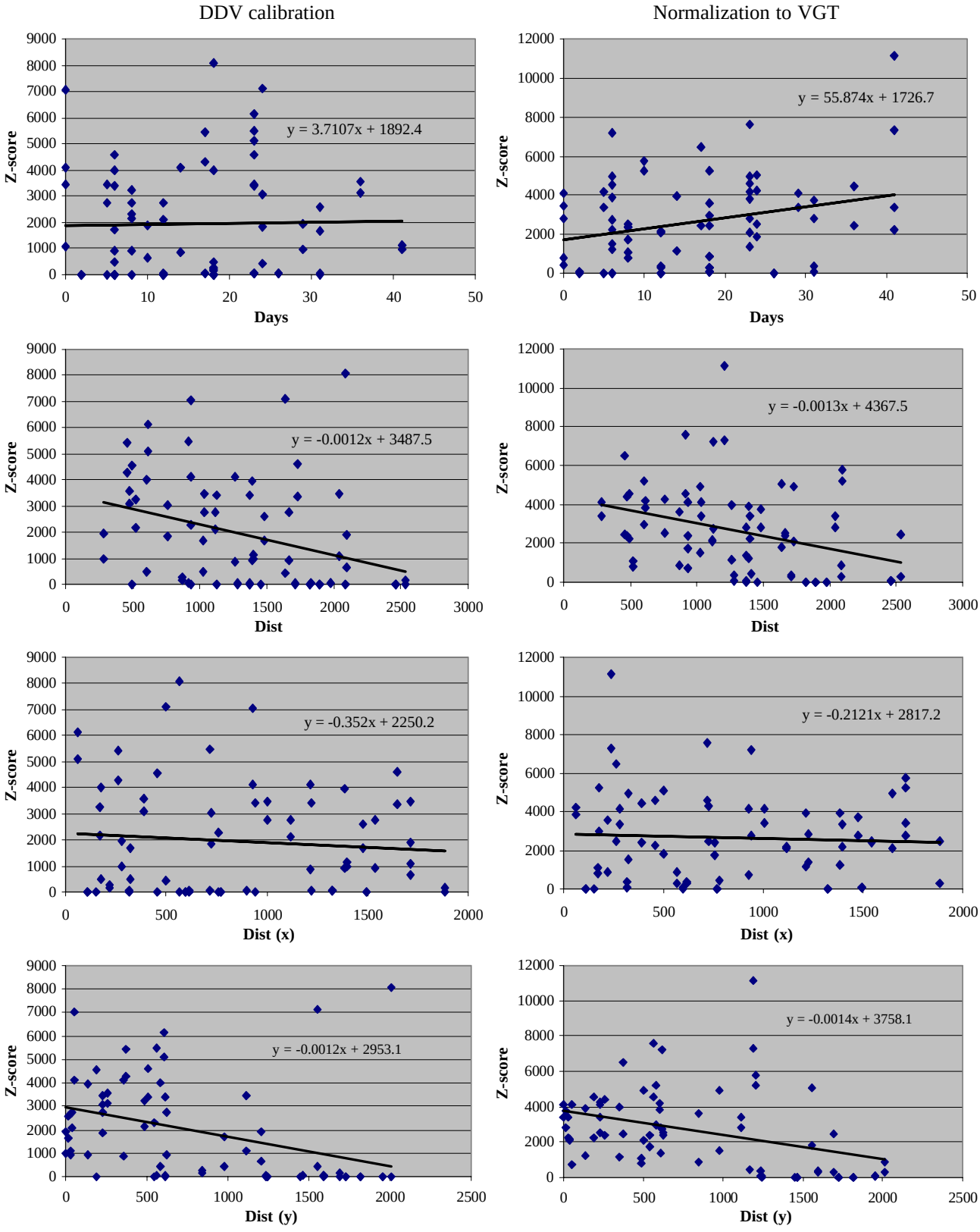

Fig. 3. Standardized Kappa statistics as a function of differences in time of day, anniversary date and distance in kilometers between scene acquisitions for scene radiometrically corrected using absolute calibration using DDV (left column) and normalization to VGT (right column).

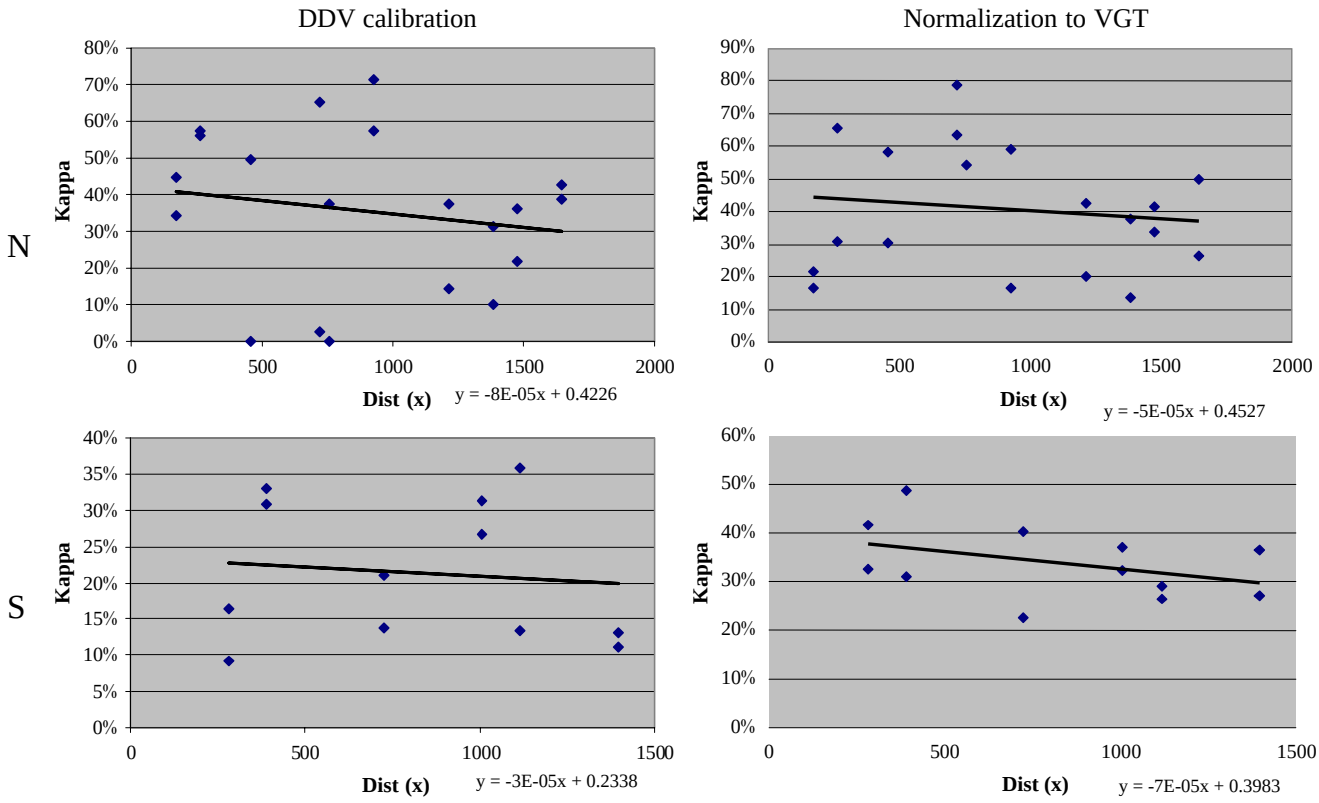

Fig. 4. Kappa statistics of northern scenes (top row) and southern scenes (bottom row) as a function of distance in kilometers between scene acquisitions for scene radiometrically corrected using absolute calibration using DDV (left column) and normalization to VGT (right column).

longitude to a distance of almost 2000 km, while calibration using DDV produced a slightly greater decrease so that nearing 2000 km, accuracy measured by the Z_{kappa} decreased to approximately 75% of accuracy at a distance of less than 200 km.

The decrease in accuracy due to Euclidean distance mentioned previously is almost entirely caused by differences in longitudinal distance between signature source and classification. The trend is such that at a distance of approximately 1200 km, classification accuracy drops to about 50% of accuracy at a distance of less than 100 km. A primary reason for this is likely the fact that the same class label is assigned to slightly different vegetation associations, and that these differences are greater in the north–south direction than in east–west. For example, needle-leaf forest in much of the boreal zone has a broadleaf understory, while further north in the transition zone and low-lying parts of the tundra, lichen and moss will often contribute to the needle-leaf signature. These two forests will both be assigned the same class label but can have different signatures.

Examination of separate landcover associations using the Kappa statistic revealed similar trends to those already reported using Z_{kappa} where the range of the factor under consideration was similar. However, a slight decrease in the Kappa statistic was noticed for both landcover associations and both radiometric correction methods with increasing longitudinal distance. This trend does not appear to be explainable by small variations in the number of classes included in the accuracy assessments and may indicate greater sensitivity of the Kappa statistic compared to Z_{kappa} when N is relatively constant.

For northern scenes, the Kappa statistic decreased by 10% for DDV and by 4% for VGT normalization to a distance of approximately 1500 km. Differences in signature extension classification accuracies between radiometric correction methods were small for northern scenes compared to southern, with normalization to VGT slightly outperforming DDV calibration. The relative magnitude of decreasing Kappa with increasing distance for the two correction methods was opposite for southern scenes, with normalization to VGT exhibiting a larger decrease of 8% Kappa compared to 3% for DDV calibration to a distance of 1200 km.

3.3. Signature extension application

An example of signature extension after normalization to VGT is presented. From Table 3, reference scene 33021 was chosen as the signature source because it contained the largest number of classes (N=26) and scene 22023 was chosen for classification because it had the largest number of classes common with 33021 (N=21). The large number of classes allowed an evaluation of the effects of class merging to simpler legends on classification accuracy. The International Geosphere–Biosphere Program (IGBP) and Interna-

Table 5
Lookup tables to merge classes from FGDC to IGBP and IPCC legends

FGDC-46 classes		IGBP-16 classes		IPCC-6 classes
1, 2, 7, 8, 9, 10	1	Evergreen Needleleaf Forest	1	Forest land
	2	Evergreen Broadleaf Forest	1	Forest land
	3	Deciduous Needleleaf Forest	1	Forest land
3, 11, 12	4	Deciduous Broadleaf Forest	1	Forest land
4, 5, 6, 13, 14, 15	5	Mixed Forest	1	Forest land
16	6	Closed Shrublands	1	Forest land
18, 19, 39, 40	7	Open Shrublands	3	Grassland
21, 22	8	Woody Savannas	1	Forest land
	9	Savannas	3	Grassland
17	10	Grasslands	3	Grassland
20	11	Permanent Wetlands	4	Wetland
32, 33, 34	12	Croplands	2	Cropland
42	13	Urban and Built-Up	5	Settlements
	14	Cropland/Natural Vegetation Mosaic	2	Cropland
45	15	Snow and Ice	6	Other land
23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 41	16	Barren or Sparsely Vegetated	6	Other land
43, 44, 46	0	Water	0	Water*

* A water class was added to the IPCC legend.

tional Panel on Climate Change (IPCC) legends were chosen for this purpose, containing 16 and 6 landcover classes, respectively. Lookup tables for conversion between legends are presented in Table 5.

Results of scene 33012 signature extension to scene 22023 using the FGDC legend are shown in Fig. 4. Similar landcover patters and colours can be seen in the reference and signature extension classifications, and the main sources of confusion are between different density classes of the same vegetation type. For example, old evergreen forest with greater than 75% cover is confused with evergreen open canopy with 40–60% cover and moss–shrub understory.

Accuracy measures for the classification in Fig. 5 are presented in Table 6 for the three legends. A water class was included in the IPCC legend in addition to the six existing landcover classes for comparison with other legends that contain a water class.

Both the overall accuracy and Kappa statistic improve as classes are merged to simpler legends to the point where at the IPCC level, the resulting classification may be considered useful. Since the improvement in classification accuracy is primarily caused by a reduced number of classes when going to simpler legends, it is not surprising that the standardized Kappa (Z_{kappa}) does not vary with the overall accuracy and Kappa statistic.

While SILC classifications may provide suitable reference data for relative comparison of radiometric correction methods for signature extension, they do not provide an absolute reference and therefore classification accuracies reported in this paper do not represent true accuracy. SILC classifications were generated using expert image interpretation, however, few were validated with ground truth data. Thus biases may exist in some of the SILC classifications caused by a different interpretation of signatures between scenes. Because the same signature might be interpreted

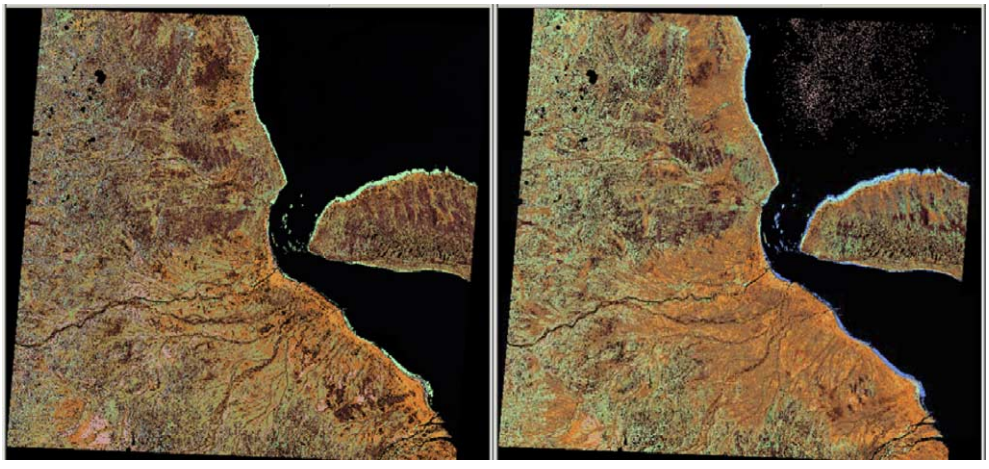

Fig. 5. Scene 22023 SILC classification using FGDC legend (left) and classification using signatures extended from scene 33012 (right) after normalization to VGT.

Table 6
Accuracy measures of scene 22023 classification using signatures extended from scene 33012 after normalization to VGT

	Overall accuracy (%)	Kappa (%)	Z_{kappa}	N
FGDC	40.7	37.2	4135	21
IGBP	61.1	48.9	4552	8
IPCC[a]	72.6	54.2	4484	5

[a] Contains a water class.

differently between regions while it actually represents the same landcover class on the ground, results from signature extension may represent real landcover more accurately than expert interpretation. More field campaigns are needed to determine absolute accuracy resulting from landcover classification using signature extension.

Landcover classification using signature extension may be envisioned as a three-stage process towards a final landcover product. The first stage uses signature extension to derive an initial clustering using spectral signatures and cluster merging to landcover. In the second stage, an experienced image interpreter can check and refine this initial classification using ECM (Beaubien et al., 1999) by splitting problem landcover classes to their original clusters and re-merging to more appropriate classes using expert knowledge. In the third stage, available ancillary reference data is incorporated to further refine the classification and provide validation data to communicate the quality of the landcover classification to the user. The signature extension approach presented in this paper is useful to get to stage one in this proposed approach, and by doing so, significantly reduces the amount of user input required to produce landcover on a scene-by-scene basis.

4. Conclusions and recommendations

There is an increasing need to produce large regional to national level landcover maps with sufficient detail to address information needs of organizations such as the IPCC. The required coverage and detail cannot be dealt with simultaneously. One solution that has been employed in large area landcover mapping is to separately classify each scene or mosaic consisting of a number of normalized scenes. This approach is feasible where sufficient resources are available including ground truth or reference data and an experienced interpreter. Given resource constraints especially in remote areas, an alternative is to extend expert knowledge embedded in an existing classification to regions beyond the classified areas via spectral signatures to reduce the resources and effort required to classify a large number of scenes.

Signature extendibility is a function of a number of factors, including image radiometry, phenology, sun-surface-sensor geometry and thematic consistency. Image radiometric correction such as absolute calibration using DDV and normalization to VGT can account for some of these factors investigated in the current study. Overall, normalization to VGT produced more stable signatures through space and time than absolute calibration using DDV. However, even after radiometric correction, above-mentioned factors continued to affect signature extendibility. Signature stability is affected by geographical distance and appears to be affected less by the difference in anniversary dates between acquisitions for peak of growing season images. An examination of the direction of signature extension revealed greater signature stability in the longitudinal direction than in latitude, likely due to greater consistency in vegetation assemblages in the east–west direction than in north south.

The example presented of the application of signature extension was chosen because of the number of classes common between the reference and classified scene and therefore may not represent the optimal case. While the utility of the classification presented at the three thematic levels considered may be limited, normalization to VGT represents a significant improvement over DDV calibration for signature extension thorough space and time.

From the findings obtained in the current study, certain recommendations can be put forth to apply signature extension for operational large area mapping with medium resolution data. First, expert classifications need to be performed to generate signatures for extension, with a greater number of classifications in the north–south direction than in the east west. The interval between adjacent classifications can be based upon the level of accuracy required and can be identified using graphs shown in Figs. 2 and 3. Second, a matrix such as the one shown in Table 3 can be useful to identify which signatures can be extended to which scenes and can allow identification of different signature sets to be applied under different conditions. Table 3 can also be useful to identify signatures to combine from different sources to form a complete set. Finally, it must be recognized that this proposed method of signature extension does not remove the interpreter from the classification process, rather it is intended to make better use of the interpreter's knowledge. A classification using extended signatures may only represent a first 'best guess', requiring further refinement by an interpreter as ancillary information on the area of interest becomes available.

Acknowledgements

The authors gratefully acknowledge Rasim Latifovic for processing the SPOT VGT data and for providing software used in the radiometric normalization method presented. We also acknowledge Richard Fernandes for his input in developing both radiometric correction methods. This paper has also benefited from comments made by three anonymous reviewers.

References

Beaubien, J., Cihlar, J., Simard, G., & Latifovic, R. (1999). Land cover from multiple Thematic Mapper scenes using a new enhancement-classification methodology. *Journal of Geophysical Research*, *104*(D22), 27909–27920.

Bezdek, J. C. (1973). *Fuzzy mathematics in pattern classification*. PhD-thesis, Applied Math Center, Cornell University, Ithaca.

Bishop, Y. M. M., Feinberg, S. E., & Hooland, P. W. (1975). *Discrete multivariate analysis theory and practice*. Cambridge, MA: MIT Press. 575 pp.

Bojinski, S., Schaepman, M., Schlaepfer, D., & Itten, K. (2003). SPECCHIO; a spectrum database for remote sensing applications. *Computers & Geosciences*, *29*, 27–38.

Butson, C., & Fernandes, R. (2004). A consistency analysis of surface reflectance and leaf area index retrieval from overlapping clear-sky Landsat ETM+imagery. *Remote Sensing of Environment*, *89*, 369–380.

Cihlar, J., Xiao, Q., Beaubien, J., Fung, K., & Latifovic, R. (1998). Classification by progressive generalization: A new automated methodology for remote sensing multichannel data. *International Journal of Remote Sensing*, *19*, 2685–2704.

Cihlar, J., Guindon, B., Beaubien, J., Latifovic, R., Peddle, D., Wulder, M., et al. (2002). From need to product: A methodology for completing a land cover map of Canada with Landsat data. *Canadian Journal of Remote Sensing*, *29*, 171–186.

Cihlar, J., Latifovic, R., Beaubien, J., Guindon, B., & Palmer, M. (2003). Thematic mapper (TM) based accuracy assessment of a land cover product for Canada derived from SPOT VEGETATION (VGT) data. *Canadian Journal of Remote Sensing*, *29*, 154–170.

Clark, R. (1990). High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research*, *95*(B8), 12653–12680.

Cohen, J. (1960). A coefficient of agreement of nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Du, Y., Cihlar, J., & Latifovic, R. (2001). Radiometric normalization, compositing, and quality control for satellite high resolution image mosaics over large areas. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 623–634.

Fedosejevs, G., O'Neill, N. T., Royer, A., Teillet, P. M., Bokoye, A. I., & McArthur, L. J. B. (2000). Aerosol optical depth for atmospheric correction of AVHRR composite data. *Canadian Journal of Remote Sensing*, *26*, 273–284.

Fernandes, R. A., & Leblanc, S. G. (in press). Appropriate linear regression techniques for the calibration of remote sensing models: When classical linear regression should not be used. *Remote Sensing of Environment*.

Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography*, *26*, 173–205.

Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B., Li, R., & Flynn, L. (1997). The MODIS 2.1-μm channel—correlation with visible reflectance for use in remote sensing of aerosols. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 1286–1297.

Latifovic, R., & Olthof, I. (2004). Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. *Remote Sensing of Environment*, *90*, 153–165.

Latifovic, R., Zhi-Liang, Z., Cihlar, J., Giri, C., & Olthof, I. (2004). Land cover mapping of North and Central America—Global Land Cover 2000. *Remote Sensing of Environment*, *89*, 116–127.

Liang, S., Fang, H., & Chen, M. (2001). Atmospheric correction of Landsat ETM+land surface imagery: Part 1. Methods. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 2490–2498.

Minter, T. C. (1978). Methods of extending crop signatures from one area to another. *Proceedings, the LACIE symposium, a technical description of the large area crop inventory experiment (LACIE), October 23–26, 1978, Houston, TX.*

Olthof, Butson, C., Fernandes, R., Fraser, R., Latifovic, R., & Orazietti, J. (submitted for publication). Landsat ETM+mosaic of northern Canada. *Canadian Journal of Remote Sensing*. 06/04.

Olthof, I., Pouliot, D., Fernandes, R., & Latifovic, R. (2005). Landsat ETM+radiometric normalization comparison for northern mapping applications. *Remote Sensing of Environment*, *95*, 388–398.

Pax-Lenney, M., Woodcock, C. E., Macomber, S. A., Gopal, S., & Song, C. (2001). Forest mapping with a generalized classifier and Landsat TM data. *Remote Sensing of Environment*, *77*, 241–250.

Vermote, E. F., Tanre, D., Deuzé, J. L., Herman, M., & Morcrette, J. J. (1997). Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 675–686.

Vogelmann, J.E, Howard, S. H., Yang, L., Larson, C. R., Wylie, B. K., & Van Driel, N. (2001). Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, *67*, 650–662.

Woodcock, C. E., Macomber, S. A., Pax-Lenney, M., & Cohen, W. B. (2001). Monitoring large areas for forest change using Landsat: Generalization across space, time and Landsat sensors. *Remote Sensing of Environment*, *78*, 194–203.

Wulder, M. A., Dechka, J. A., Gillis, M. A., Luther, J. E., Hall, R. J., Beaudoin, A., et al. (2003). Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *Forestry Chronicle*, *79*, 1075–1083.